

日本国特許庁

PATENT OFFICE  
JAPANESE GOVERNMENT

1C875 U.S. PTO

09/651058



別紙添付の書類に記載されている事項は下記の出願書類に記載されて  
る事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed  
this Office.

願年月日  
Date of Application:

1999年 8月31日

願番号  
Application Number:

平成11年特許願第246393号

願人  
Applicant(s):

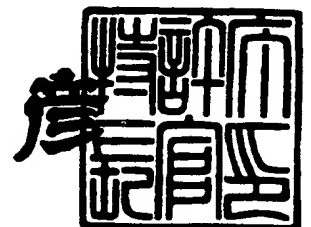
パイオニア株式会社

CERTIFIED COPY OF  
PRIORITY DOCUMENT

2000年 4月21日

特許庁長官  
Commissioner,  
Patent Office

近藤隆彦



出証番号 出証特2000-3030183

【書類名】 特許願

【整理番号】 54P0080

【提出日】 平成11年 8月31日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 7/08  
G10L 9/18

【発明者】

【住所又は居所】 埼玉県川越市山田字西町 2 5 番地 1 パイオニア株式会社  
社川越工場内

【氏名】 依田 正太郎

【特許出願人】

【識別番号】 000005016

【氏名又は名称】 パイオニア株式会社

【代理人】

【識別番号】 100063565

【弁理士】

【氏名又は名称】 小橋 信淳

【手数料の表示】

【予納台帳番号】 011659

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声認識システム

【特許請求の範囲】

【請求項 1】 複数個の音検知手段と、

前記複数個の音検知手段から出力される各入力信号の中から音声認識に適した音声信号を決定する決定手段と、

前記決定手段で決定された音声信号に基づいて音声認識を行う音声認識手段とを具備することを特徴とする音声認識システム。

【請求項 2】 前記決定手段は、前記複数個の音検知手段から出力される各入力信号の平均  $S/N$  値と平均音声パワーとを求め、前記平均  $S/N$  値と平均音声パワーがそれぞれ所定の閾値より大きい値となった入力信号を、前記音声認識に適した音声信号と決定することを特徴とする請求項 1 に記載の音声認識システム。

【請求項 3】 前記決定手段は、前記平均  $S/N$  値と平均音声パワーがそれぞれ所定の閾値より大きな値となった音声信号のうち、平均  $S/N$  値と平均音声パワーの値の大きさに従って、音声認識に適した音声信号の候補順位を決定し、

前記音声認識手段は、前記候補順位の上位の音声信号に基づいて順次に音声認識を行うことを特徴とする請求項 2 に記載の音声認識システム。

【請求項 4】 前記決定手段は、前記音声認識に適した音声信号以外の入力信号を雑音信号とすることを特徴する請求項 1 ～ 3 のいずれか 1 項に記載の音声認識システム。

【請求項 5】 前記決定手段は、前記音声認識に適した音声信号以外の入力信号のうち、平均  $S/N$  値と平均音声パワーとが最も小さな値となる入力信号を雑音信号とすることを特徴とする請求項 1 ～ 4 のいずれか 1 項に記載の音声認識システム。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

本発明は、発話音声によって電子機器の制御又は操作等を可能にする音声認識

システムに関するものである。

【 0 0 0 2 】

【従来の技術】

従来、この種の音声認識システムとして、車載用オーディオシステムや車載用ナビゲーションシステム等の電子機器に適用されたものが知られている。

【 0 0 0 3 】

音声認識システムを備えた車載用オーディオシステムでは、例えば、搭乗者が所望の放送局の局名を発話すると、音声認識システムが発話音声を音声認識し、認識結果に基づいてラジオ受信機の受信周波数を自動選局させることで、搭乗者の利便性の向上が図られている。

【 0 0 0 4 】

また、車載用オーディオシステムに備えられているMD (Mini Disc) プレーヤーやCD (Compact Disc) プレーヤー等に、搭乗者が記録再生媒体を挿入し、その記録再生媒体に記録されている楽曲の名称を発話すると、音声認識システムが発話音声を音声認識して楽曲を自動的に再生させる等の機能も備えられている。

【 0 0 0 5 】

車載用ナビゲーションシステムでは、運転手等が行きたい目的地の名称を発話すると、音声認識システムが発話音声を音声認識して、現在地から目的地までの走行経路を地図表示させる等の機能が備えられ、この結果、運転手は運転操作に集中することができ安全運転の確保が図れる等の効果が得られるようになっている。

【 0 0 0 6 】

【発明が解決しようとする課題】

ところで、上記従来の音声認識システムは、一人の発話者を対象とするものであったため、発話音声を入力するためのマイクロフォンが使用頻度の高い運転手に最も近い位置に1個だけ設けられていた。

【 0 0 0 7 】

このため、例えば、マイクロフォンから遠い位置に着座している他の搭乗者は

、マイクロフォンの音声入力レベルを十分に確保すべく、マイクロフォンに向かって大声で発話しなければならないという問題があった。また、音声認識システムの音声認識率を高めるためには、他の搭乗者は発話音声車を車内雑音の影響を受けることなくマイクロフォンに入力させるべく、マイクロフォンに向かって大声で発話しなければならないという問題があった。

【0008】

本発明はこのような問題点に鑑みて成されたものであり、より利便性の高い音声認識システムを提供することを目的とする。

【0009】

【課題を解決するための手段】

上記目的を達成するための本発明は、複数個の音検知手段と、上記複数個の音検知手段から出力される各入力信号の中から音声認識に適した音声信号を決定する決定手段と、上記決定手段で決定された音声信号に基づいて音声認識を行う音声認識手段とを具備することを特徴とする。

【0010】

また、上記決定手段は、上記複数個の音検知手段から出力される各入力信号の平均 S/N 値と平均音声パワーとを求め、上記平均 S/N 値と平均音声パワーがそれぞれ所定の閾値より大きい値となった入力信号を上記音声認識に適した音声信号と決定することを特徴とする。

【0011】

上記決定手段は、上記平均 S/N 値と平均音声パワーがそれぞれ所定の閾値より大きな値となった音声信号のうち、平均 S/N 値と平均音声パワーの値の大きさに従って音声認識に適した音声信号の候補順位を決定し、上記音声認識手段は、上記候補順位の上位の音声信号に基づいて順次に音声認識を行うことを特徴とする。

【0012】

また、上記決定手段は、上記音声認識に適した音声信号以外の入力信号を雑音信号とすることを特徴とする。

## 【0013】

また、上記決定手段は、上記音声認識に適した音声信号以外の入力信号のうち、平均S/N値と平均音声パワーとが最も小さな値となる入力信号を雑音信号とすることを特徴とする。

## 【0014】

これらの構成によれば、発話者が所望の発話をする、と、複数の音検知手段から出力される各入力信号の中から、音声認識を行うのに適した音声信号と雑音信号が自動的に決定され、その決定された音声信号と雑音信号に基づいて音声認識が行われる。この結果、発話者は特定の音検知手段に対して発話するという意識を持つことなく、発話するだけで適切な音声認識がなされることとなり、利便性の向上等が実現される。

## 【0015】

## 【発明の実施の形態】

以下、本発明の実施の形態を図面を参照して説明する。尚、一実施形態として、車載用オーディオシステムや車載用ナビゲーションシステム等の車載用電子機器を発話音声によって制御又は操作することを可能にする音声認識システムについて説明する。

## 【0016】

図1は、本実施形態に係る音声認識システムの構成を示すブロック図である。同図において、本音声認識システムは、音検知手段としての複数のマイクロフォン $M_1 \sim M_N$ 、複数の前置回路 $CC_1 \sim CC_N$ 、マルチプレクサ1、A/D変換器2、デマルチプレクサ3、記憶部4、発話音声検出部5、データ解析部6、音声認識部7、制御部8、発話スイッチ9を備えて構成されている。

## 【0017】

ここで、複数の前置回路 $CC_1 \sim CC_N$ 、マルチプレクサ1、A/D変換器2、デマルチプレクサ3、記憶部4、発話音声検出部5、データ解析部6、制御部8、は、音声認識を行うのに適した音声信号と雑音信号とを決定するための決定手段を構成している。

## 【 0 0 1 8 】

発話スイッチ 9 は 1 個だけ備えられ、運転手の座席近傍、例えば車両のフロントダッシュボードや前部ドアの一端に設けられる。

## 【 0 0 1 9 】

制御部 8 は、マイクロプロセッサ (MPU) を有し、発話スイッチ 9 がオン操作されてオン信号 SW が供給されると、マイクロフォン  $M_1 \sim M_N$  に集音動作を開始させると共に、本音声認識システム全体の動作を制御する。尚、詳細については動作説明と共に述べるが、発話音声検出部 5 には、どのマイクロフォンに発話が行なわれたかを判断するための連続回数カウンタ  $FC_1 \sim FC_N$  が備えられている。

## 【 0 0 2 0 】

各マイクロフォン  $M_1 \sim M_N$  は、各搭乗者の発話音声を集音し易い位置、例えば各搭乗者の座席近傍に設けられる。

## 【 0 0 2 1 】

一例として 4 個のマイクロフォン  $M_1 \sim M_4$  を車室内に設置する場合には、図 2 (a) の平面図に示すように、マイクロフォン  $M_1, M_2$  を運転席と助手席の前方にそれぞれ設置し、マイクロフォン  $M_3, M_4$  を後部座席の前方、例えば天井部分や、運転席と助手席の背部等に設置することで、各マイクロフォン  $M_1 \sim M_4$  と各搭乗者とを対応付けるようにする。

## 【 0 0 2 2 】

また、他の一例として、図 2 (b) の平面図に示すように、マイクロフォン  $M_1, M_2$  を運転席側の前部ドアと助手席側の前部ドアにそれぞれ設置し、マイクロフォン  $M_3, M_4$  を各後部座席側の後部ドア等に設置することにより、各マイクロフォン  $M_1 \sim M_4$  と各搭乗者とを対応付けるようにする。

## 【 0 0 2 3 】

また、更に他の一例として、図 2 (a) (b) を組み合わせた位置に各マイクロフォン  $M_1 \sim M_4$  を設置してもよい。つまり、マイクロフォン  $M_1$  を図 2 (a) に示す運転席の前方に設置する場合には図 2 (b) に示す運転席側の前部ドアには設置せず、マイクロフォン  $M_1$  を図 2 (b) に示す運転席側の前部ドアに設置

する場合には図 2 (a) に示す運転席の前方には設置しないように、図 2 (a) (b) に示す何れか一方の位置を選択して、運転席に着座する運転手に対して 1 個のマイクロフォン  $M_1$  を設置するようにする。他のマイクロフォン  $M_2 \sim M_4$  についても同様に、図 2 (a) (b) に示す何れか一方の位置を選択して、それぞれ 1 個ずつ設置する。

## 【 0 0 2 4 】

また、座席数の多いワゴン車等の場合には、図 3 (a) (b) の平面図に示すように、座席数に応じてより多くのマイクロフォン  $M_1 \sim M_6$  等を設置し、これらのマイクロフォン  $M_1 \sim M_6$  も各搭乗者の発話音声を集音し易い位置に 1 個ずつ設ける。また、図 3 (a) (b) を組み合わせた位置に各マイクロフォン  $M_1 \sim M_4$  を 1 個ずつ設置してもよい。

## 【 0 0 2 5 】

尚、上述したマイクロフォンの各配置例はあくまでも一例である。実際には、本発明のシステムで使用するシステム情報は、予め搭乗者からマイクロフォンまでの音響伝達特性を考慮して構築される。そのため、マイクロフォンの設置条件が厳密に制限されることはない。また、マイクロフォンの数は、車種等に応じて予め決められている最大搭乗者数と同数又はそれより少ない数に決めることができる。

## 【 0 0 2 6 】

また、単に各マイクロフォンから各搭乗者までの距離を等しくするというのではなく、車室内の音響特性を実験等によって予め解析しておき、この解析結果に基づいて、各マイクロフォンから各搭乗者までのそれぞれの音響伝達特性がほぼ等しくなるように、それぞれの距離と設置場所が決めることができる。

## 【 0 0 2 7 】

再び図 1 において、各マイクロフォン  $M_1 \sim M_N$  には、前置回路  $CC_1 \sim CC_N$  が対応付けて接続され、これによって  $N$  チャンネル分の信号処理系統が構成されている。

## 【 0 0 2 8 】

各前置回路  $CC_1 \sim CC_N$  には、各マイクロフォン  $M_1 \sim M_N$  から供給される入力



信号  $S_1 \sim S_N$  を信号処理可能な振幅レベルに増幅する増幅器（図示省略）と、増幅器で増幅された入力信号のうち所定周波数成分のみを通過させるバンドパスフィルタ（図示省略）が備えられており、各バンドパスフィルタを通過した入力信号  $S_1' \sim S_N'$  をマルチプレクサ 1 に供給するようになっている。

【 0 0 2 9 】

上記の各バンドパスフィルタは、入力信号  $S_1 \sim S_N$  に含まれる低域雑音を除去するための低域カットオフ周波数  $f_L$ （例えば、 $f_L = 100 \text{ Hz}$ ）と、ナイキスト周波数を考慮した高域カットオフ周波数  $f_H$  が設定されており、これら低域カットオフ周波数  $f_L$  と高域カットオフ周波数  $f_H$  との周波数範囲内に、人間が発話する音声の音声周波数帯域が含まれるように設定されている。

【 0 0 3 0 】

マルチプレクサ 1 は、図 4 に示すように、 $N$  チャンネル分のアナログスイッチ  $AS_1 \sim AS_N$  を備えて構成されている。各アナログスイッチ  $AS_1 \sim AS_N$  の入力端に各前置回路  $CC_1 \sim CC_N$  からの入力信号  $S_1' \sim S_N'$  が供給され、各アナログスイッチ  $AS_1 \sim AS_N$  の出力端は共通接続されて A/D 変換器 2 に接続されている。そして、制御部 8 から供給されるチャンネル切換え信号  $CH_1 \sim CH_N$  に応じて、入力信号  $S_1' \sim S_N'$  を排他的に切換えて A/D 変換器 2 に供給する。

【 0 0 3 1 】

A/D 変換器 2 は、マルチプレクサ 1 から順次に供給される入力信号  $S_1' \sim S_N'$  を所定のサンプリング周波数  $f$  に同期してそれぞれデジタルの入力データ  $D_1 \sim D_N$  に変換し、デマルチプレクサ 3 に供給する。

【 0 0 3 2 】

上記のサンプリング周波数  $f$  は、制御部 8 からのサンプリングクロック  $CK_{AD}$  によって設定されると共に、アンチエイリアシングを考慮した周波数に決められている。より具体的には、サンプリング周波数  $f$  は、前置回路  $CC_1 \sim CC_N$  に備えられている上記バンドパスフィルタの高域カットオフ周波数  $f_H$  の約 2 倍以上に決められ、例えば、 $8 \text{ kHz}$  ないし  $11 \text{ kHz}$  の範囲内の周波数に決められている。

## 【 0 0 3 3 】

デマルチプレクサ 3 は、図 4 に示すように、 $N$  チャンネル分のアナログスイッチ  $AW_1 \sim AW_N$  を備えて構成されている。各アナログスイッチ  $AW_1 \sim AW_N$  の入力端は共通接続されて A/D 変換器 2 の出力端に接続され、各アナログスイッチ  $AW_1 \sim AW_N$  の各出力端は、記憶部 4 に備えられている  $N$  チャンネル分の記憶領域  $ME_1 \sim ME_N$  に接続されている。そして、制御部 8 から供給されるチャンネル切換え信号  $CH_1 \sim CH_N$  に応じて、入力データ  $D_1 \sim D_N$  を排他的に切換えて各記憶領域  $ME_1 \sim ME_N$  に供給する。

## 【 0 0 3 4 】

ここで、図 5 のタイミングチャートを参照して、マルチプレクサ 1 と A/D 変換器 2 及びデマルチプレクサ 3 の動作を説明する。同図において、発話スイッチ 9 がオン操作されると、それによって生じるオン信号  $SW$  を制御部 8 が受信し、サンプリングクロック  $CK_{ADC}$  とチャンネル切換え信号  $CH_1 \sim CH_N$  を出力する。

## 【 0 0 3 5 】

サンプリングクロック  $CK_{ADC}$  は、サンプリング周波数  $f$  の逆数  $1/f$  の周期 (サンプリング周期)  $T$  の間に  $N$  回の論理反転を繰り返すパルス波形であり、各チャンネル切換え信号  $CH_1 \sim CH_N$  は、サンプリングクロック  $CK_{ADC}$  の 1 周期  $T/N$  毎に順番に論理 “1” となるパルス波形である。

## 【 0 0 3 6 】

マルチプレクサ 1 は、チャンネル切換え信号  $CH_1 \sim CH_N$  が順番に論理 “1” となる周期  $T/N$  に同期して排他的に導通/非導通の切換え動作を行う。このため、入力信号  $S_1' \sim S_N'$  は、周期  $T/N$  に同期して順番に A/D 変換器 2 に供給されてデジタルデータ  $D_1 \sim D_N$  に変換される。一方、デマルチプレクサ 3 も、チャンネル切換え信号  $CH_1 \sim CH_N$  が順番に論理 “1” となる周期  $T/N$  に同期して排他的に導通/非導通の切換え動作を行う。このため、A/D 変換器 2 から出力される入力データ  $D_1 \sim D_N$  は、周期  $T/N$  に同期して記憶領域  $ME_1 \sim ME_N$  に振り分けられて記憶される。

## 【 0 0 3 7 】

このように、サンプリング周期  $T (= 1 / f)$  の間に  $N$  チャンネル分の入力信号  $S_1' \sim S_N'$  をサンプリングしてこれを繰り返すことで、1 個の  $A/D$  変換器 2 でも、 $N$  チャンネル分の入力データ  $D_1 \sim D_N$  をサンプリング周波数  $f$  に同期して生成し、所定の記憶領域  $ME_1 \sim ME_N$  に振り分けて記憶させることが可能となっている。

## 【 0 0 3 8 】

記憶部 4 は、半導体メモリで形成され、上述した  $N$  チャンネル分の記憶領域  $ME_1 \sim ME_N$  を備えている。すなわち、各記憶領域  $ME_1 \sim ME_N$  は、各マイクロフォン  $M_1 \sim M_N$  に対応付けて備えられている。

## 【 0 0 3 9 】

更に図 4 に示すように、各記憶領域  $ME_1 \sim ME_N$  には、それぞれの入力データ  $D_1 \sim D_N$  を所定サンプル数のフレーム単位で記憶するために、複数のフレーム領域  $MF_1, MF_2 \dots$  が備えられている。

## 【 0 0 4 0 】

記憶領域  $ME_1$  を代表して述べれば、各フレーム領域  $MF_1, MF_2 \dots$  は、デマルチプレクサ 3 から供給される入力データ  $D_1$  を、制御部 8 からのアドレス信号  $ADR_1$  に従って予め決められたサンプル数（本実施形態では、256 サンプル）ずつ順番に記憶するようになっている。つまり、図 5 に示した期間（256  $\times T$ ）をフレーム期間  $TF$  と呼び、この 1 フレーム期間（1  $TF$ ）毎に、各フレーム領域  $MF_1, MF_2 \dots$  に 256 個ずつの入力データ  $D_1$  を記憶させるようになっている。また、各フレーム領域  $MF_1, MF_2 \dots$  に記憶される 1 フレーム期間（1  $TF$ ）分の入力データをフレームデータと呼んでいる。

## 【 0 0 4 1 】

そして、残余の記憶領域  $ME_2 \sim ME_N$  に備えられている各フレーム領域  $MF_1, MF_2 \dots$  にも、それぞれの入力データ  $D_2 \sim D_N$  を 1 フレーム期間（1  $TF$ ）毎に、256 サンプルずつを記憶させるようになっている。

## 【 0 0 4 2 】

発話音声検出部 5 とデータ解析部 6 は、DSP (Digital Signal Processor)

で形成されている。

【 0 0 4 3 】

発話音声検出部 5 は、各記憶領域  $ME_1 \sim ME_N$  の各フレーム領域  $MF_1, MF_2 \dots$  にフレームデータが記憶される度に、最新のフレームデータの LPC 残差を演算し、その演算値が所定の閾値  $THD_1$  以上となるか判定する。演算値が所定の閾値  $THD_1$  以上となった場合には、最新のフレームデータを発話によって生じた音声フレームデータであると判定し、演算値が所定の閾値  $THD_1$  未満の場合には、最新のフレームデータを発話によらない入力データ、すなわち車室内の雑音によって生じた雑音フレームデータであると判定する。

【 0 0 4 4 】

更に発話音声検出部 5 は、LPC 残差の演算値が 3 フレーム期間 ( $3TF$ ) に亘って連続して閾値  $THD_1$  以上となった場合に、その 3 フレーム期間 ( $3TF$ ) のフレームデータを明らかに発話によって生じた音声フレームデータであると確定し、その確定結果を示す音声検出データ  $DCT_1$  を制御部 8 へ転送する。

【 0 0 4 5 】

より詳細には、各記憶領域  $ME_1 \sim ME_N$  の各フレーム領域  $MF_1, MF_2 \dots$  に記憶されるフレームデータの LPC 残差を各チャンネル毎に個別演算し、更に、各チャンネル毎に求められた LPC 残差の演算値を閾値  $THD_1$  と比較することで、各チャンネル毎に上記の発話による音声フレームデータか否かの判定を行うようになっている。

【 0 0 4 6 】

つまり、マイクロフォン  $M_1$  に関連する第 1 チャンネルの LPC 残差の演算値が  $\varepsilon_1$ 、マイクロフォン  $M_2$  に関連する第 2 チャンネルの LPC 残差の演算値が  $\varepsilon_2$ 、以下同様にマイクロフォン  $M_3 \sim M_N$  に関連する第 3 ～ 第  $N$  チャンネルの各 LPC 残差の演算値が  $\varepsilon_3 \sim \varepsilon_N$  であった場合には、各々の演算値  $\varepsilon_1 \sim \varepsilon_N$  を閾値  $THD_1$  と比較する。そして、閾値  $THD_1$  以上となったチャンネルに該当するフレームデータを発話によって生じた音声フレームデータであると判定し、更に、3 フレーム期間 ( $3TF$ ) に亘って連続して LPC 残差の演算値が閾値  $THD_1$  以上となったチャンネルの音声フレームデータを、明らかに発話によって生じた

音声フレームデータであると確定することとしている。

【 0 0 4 7 】

例えばマイクロフォン $M_1$ に向かって発話が行われ、残余のマイクロフォン $M_2 \sim M_N$ には発話音声が入力されなかったような場合には、マイクロフォン $M_1$ に関連するチャンネルの記憶領域 $ME_1$ に記憶されたフレームデータだけが発話によって生じた音声フレームデータであると判定及び確定され、残余のマイクロフォン $M_2 \sim M_N$ に関連するチャンネルの記憶領域 $ME_2 \sim ME_N$ に記憶されたフレームデータは、車室内の雑音によって生じた雑音フレームデータであると判定される。

【 0 0 4 8 】

また、例えばマイクロフォン $M_1$ に向かって発話が行われると共に、マイクロフォン $M_2$ にはその発話音声が入り込んで入力され、残余のマイクロフォン $M_3 \sim M_N$ には発話音声が入力されなかったような場合には、マイクロフォン $M_1$ と $M_2$ に関連するチャンネルの記憶領域 $ME_1$ と $ME_2$ に記憶された両方のフレームデータが発話によって生じた音声フレームデータであると判定及び確定され、残余のマイクロフォン $M_3 \sim M_N$ に関連するチャンネルの記憶領域 $ME_3 \sim ME_N$ に記憶されたフレームデータは、雑音フレームデータであると判定される。

【 0 0 4 9 】

このように、発話音声検出部 5 は、記憶領域 $ME_1 \sim ME_N$ に記憶されたフレームデータの各 L P C 残差を演算して閾値 T H D 1 と比較することで、何れのマイクロフォンに発話音声が入力されたかの判定と、発話音声が入力されたフレーム期間の判定を行い、これらの判定情報を有する音声検出データ D C T 1 を制御部 8 へ転送する。

【 0 0 5 0 】

尚、音声検出データ D C T 1 は、上記の 3 フレーム期間以上連続して音声フレームデータが記憶された記憶領域（以下、音声記憶チャンネルという）とそのフレーム領域（以下、音声記憶フレームという）とを示す所定のコードデータとして制御部 8 へ転送される。

【 0 0 5 1 】

具体例として、音声検出データ D C T 1 は、 $D C T 1 \{CH_1 (TF_1, TF_2 \sim TF_m), CH_2 (TF_1, TF_2 \sim TF_m), \dots, CH_N (TF_1, TF_2 \sim TF_m)\}$  等の一般形式のデータ構造を有している。 $CH_1 \sim CH_N$  は各チャンネルを表すフラグデータ、 $TF_1, TF_2 \sim TF_m$  は、各フレーム領域  $MF_1, MF_2 \sim MF_m$  に対応するフラグデータとなっている。

【 0 0 5 2 】

そして、例えばマイクロフォン  $M_1$  のみに発話音声が入力され、且つ第 3 番目以降のフレーム領域  $MF_3, MF_4 \dots$  に音声フレームデータが記憶されたような場合は、 $D C T 1 \{1 (0, 0, 1, 1, \sim 1), 0 (0, 0, 0 \sim 0), \dots, 0 (0, 0, 0 \sim 0)\}$  となるバイナリーコードの音声検出データ D C T が制御部 8 へ転送される。

【 0 0 5 3 】

この音声検出データ D C T 1 が制御部 8 へ転送されると、制御部 8 は音声検出データ D C T 1 に基づいて、音声記憶チャンネルと音声記憶フレームを示す制御データ C N T 1 を生成して、データ解析部 6 に供給する。

【 0 0 5 4 】

データ解析部 6 は、最適音声決定部 6 a、雑音決定部 6 b、平均 S / N 値演算部 6 c、平均音声パワー演算部 6 d、平均雑音パワー演算部 6 e、音声条件テーブル 6 f、雑音選択テーブル 6 g を備えて構成されており、制御部 8 からの制御データ C N T 1 が供給されると、音声認識に適した音声フレームデータと雑音フレームデータを決定するための処理を開始する。

【 0 0 5 5 】

ここで、平均音声パワー演算部 6 d は、制御データ C N T 1 から音声記憶チャンネル及び音声記憶フレームの情報を取得し、記憶領域  $ME_1 \sim ME_N$  のうちのこれら音声記憶チャンネル及び音声記憶フレームに該当する記憶領域から音声フレームデータを読み出して、各チャンネル毎に、音声フレームデータの平均音声パワー  $P(n)$  を演算する。尚、平均音声パワー  $P(n)$  の変数  $n$  はチャンネル番号を示す変数である。

## 【0056】

例えば、図6(a)～(d)に示すように、チャンネル $CH_1 \sim CH_4$ に当たる記憶領域 $ME_1 \sim ME_4$ に音声フレームデータが記憶されている場合には、各チャンネル毎に、発話が開始された時点 $t_s$ から予め決められた複数フレーム期間( $m_2 \times TF$ )内に該当する複数の音声フレームデータの平均音声パワー $P(1) \sim P(4)$ を演算する。尚、平均音声パワー $P(n)$ は、フレーム期間( $m_2 \times TF$ )内の音声フレームデータを2乗加してフレーム期間( $m_2 \times TF$ )で除算することにより演算される。

## 【0057】

平均雑音パワー演算部6eは、制御データ $CNT1$ から音声記憶チャンネル及び音声記憶フレームの情報を取得し、記憶領域 $ME_1 \sim ME_N$ のうちのこれら音声記憶チャンネルに該当する記憶領域から、音声フレームデータよりも複数フレーム期間( $m_1 \times TF$ )前の雑音フレームデータを読み出し、各チャンネル毎に、雑音フレームデータの平均雑音パワー $NP(n)$ を演算する。尚、 $NP(n)$ の変数 $n$ は音声チャンネルを示す変数であり、また、平均雑音パワー $NP(n)$ は、フレーム期間( $m_1 \times TF$ )内の雑音フレームデータを2乗加算してフレーム期間( $m_1 \times TF$ )で除算することにより演算される。

## 【0058】

例えば、図6(a)～(d)に示すように、チャンネル $CH_1 \sim CH_4$ に当たる記憶領域 $ME_1 \sim ME_4$ に音声フレームデータが記憶されている場合には、発話が開始された時点 $t_s$ (音声フレームデータの記憶開始時点)よりも複数フレーム期間( $m_1 \times TF$ )前に該当する複数の雑音フレームデータの平均雑音パワー $NP(n)$ が演算される。

## 【0059】

平均 $S/N$ 値演算部6cは、平均音声パワー演算部6dで演算された平均音声パワー $P(n)$ と平均雑音パワー演算部6eで演算された平均雑音パワー $NP(n)$ に基づいて、音声チャンネル毎に、信号対雑音比の値を表す $S/N$ 値 $SN(n)$ を演算する。

## 【0060】

例えば、図6(a)～(d)に示すように、チャンネル $CH_1 \sim CH_4$ が音声チャンネルであった場合には、次式(1)～(4)に基づいて、各チャンネル $CH_1 \sim CH_4$ の $S/N$ 値 $SN(1) \sim SN(4)$ を演算する。

## 【0061】

$$SN(1) = P(1) / NP(1) \quad \cdots (1)$$

$$SN(2) = P(2) / NP(2) \quad \cdots (2)$$

$$SN(3) = P(3) / NP(3) \quad \cdots (3)$$

$$SN(4) = P(4) / NP(4) \quad \cdots (4)$$

また、上記式(1)～(4)によって演算された $S/N$ 値 $SN(1) \sim SN(4)$ のそれぞれの対数値を、各チャンネル $CH_1 \sim CH_4$ の $S/N$ 値 $SN(1) \sim SN(4)$ としてもよい。

## 【0062】

最適音声決定部6aは、平均 $S/N$ 値演算部6cで求められた平均 $S/N$ 値 $SN(n)$ を所定の閾値 $THD2$ と比較すると共に、平均音声パワー演算部6dで求められた平均音声パワー $P(n)$ を所定の閾値 $THD3$ と比較し、これらの比較結果を図7に示す音声条件テーブル6fと対比することで、何れのチャンネルの音声フレームデータが音声認識処理に適用するのに好適か判定する。

## 【0063】

つまり、音声条件テーブル6fには、図7に示すように、平均 $S/N$ 値と閾値 $THD2$ との関係と平均音声パワーと閾値 $THD3$ との関係とによって音声フレームデータを等級(ranking)付けるための参照データが記憶されており、最適音声決定部6aは、上記の比較結果に基づいて音声条件テーブル6fを参照することで、音声認識処理に適した音声フレームデータを等級付けし、最も上位の等級の音声フレームデータを音声認識処理に最も適していると判定する。

## 【0064】

具体的には、平均 $S/N$ 値が閾値 $THD2$ 以上で且つ平均パワー値が閾値 $THD3$ 以上となった音声フレームデータを等級1(Rnk1)、平均 $S/N$ 値が閾値 $THD2$ 以上で且つ平均パワー値が閾値 $THD3$ 未満となった音声フレームデータ



を等級 2 (Rnk2)、平均 S/N 値が閾値 THD 2 未満で且つ平均パワー値が閾値 THD 3 以上となった音声フレームデータを等級 3 (Rnk3)、平均 S/N 値が閾値 THD 2 未満で且つ平均パワー値が閾値 THD 3 未満となった音声フレームデータを等級 4 (Rnk4) と判定する。

【 0 0 6 5 】

更に、全てのチャンネルの音声フレームデータの中で、平均 S/N 値と平均パワー値が最大となった音声フレームデータを等級 0 (Rnk0) と判定する。

【 0 0 6 6 】

そして、等級 0 (Rnk0) となった音声フレームデータを音声認識処理に最も適した候補 (第 1 候補) と判定する。更に、等級 1 (Rnk1) となった音声フレームデータを音声認識処理に適した次の候補 (第 2 候補) と判定する。更にまた、等級 1 (Rnk1) となったチャンネルが複数存在した場合には、平均 S/N 値及び平均パワー値の大きなものほど上位の候補と判定する。

【 0 0 6 7 】

更に、等級 2 (Rnk2) ~ 等級 4 (Rnk4) に該当する音声フレームデータを音声認識処理に適さないものとして音声認識処理の対象外とする。

【 0 0 6 8 】

このように、最適音声決定部 6 a は、平均 S/N 値  $SN(n)$  と平均音声パワー  $P(n)$  をそれぞれ閾値 THD 2, THD 3 と比較し、これらの比較結果を図 7 に示す音声条件テーブル 6 f と対比することで、音声認識処理に適した音声フレームデータを判定し、更に、音声認識処理に適した音声フレームデータに優先順位を付ける。そして、優先順位を示す音声候補データ DCT 2 を制御部 8 へ転送する。

【 0 0 6 9 】

雑音決定部 6 b は、最適音声決定部 6 a で求められた N チャンネル分の全等級の組み合わせを図 8 に示す雑音選択テーブル 6 g と対比し、組み合わせが一致した場合のチャンネルを雑音のチャンネルと判定する。

【 0 0 7 0 】

例えば、図 8 中の「場合 1」に示すように、最適音声決定部 6 a で求められた

第 1 チャンネル  $CH_1$  以降の各チャンネルの等級が  $(Rnk0)$  ,  $(Rnk1)$  ,  $(Rnk2)$  ,  $(Rnk1)$  , ……であった場合には、雑音のチャンネルを第 3 チャンネル  $CH_3$  と判定する。そして、雑音のチャンネルを示す雑音候補データ  $DCT3$  を制御部 8 へ転送する。

## 【 0 0 7 1 】

したがって、最適音声決定部 6 a で音声認識に適した音声フレームデータの候補が判定されると、雑音決定部 6 b は、図 8 中の各「場合」を参照することにより、その音声認識に適した音声フレームデータの候補に対応する雑音チャンネルを判定する。この結果、音声認識に適した音声フレームデータの候補と、雑音を集音したマイクロフォンによって得られた雑音データとが関連付けて判定される。

## 【 0 0 7 2 】

尚、図 8 に示す雑音選択テーブル 6 g の各場合 1 , 2 , 3 …は、全てのマイクロフォン  $M_1 \sim M_N$  を車室内に実際に設置して、実際に搭乗者が様々な位置で発話したときに得られる音響特性の実験結果に基づいて予め設定されている。

## 【 0 0 7 3 】

このように音声候補データ  $DCT2$  と雑音候補データ  $DCT3$  が制御部 8 に供給されると、制御部 8 は、記憶領域  $ME_1 \sim ME_N$  のうち、音声候補データ  $DCT2$  に基づいて第 1 候補のチャンネルに該当する記憶領域をメモリアクセスし、最も音声認識処理に適した音声フレームデータを読み出して音声認識部 7 へ供給する。更に、雑音候補データ  $DCT3$  に基づいて雑音のチャンネルに該当する記憶領域をメモリアクセスし、雑音フレームデータを読み出して音声認識部 7 へ供給する。

## 【 0 0 7 4 】

音声認識部 7 は、記憶部 4 により供給される音声フレームデータと雑音フレームデータに基づいて  $SS$  (Spectrum Subtraction) 処理、エコーキャンセラー、ノイズキャンセラー、CMN 処理等の周知の処理を行うことにより、音声フレームデータ中に含まれている雑音成分を除去し、その雑音成分の除去された音声フレームデータに基づいて音声認識処理を行い、更に、音声認識結果のデータ  $Dou$

tを出力する。

【0075】

尚、音声認識部7が上記の音声認識処理に適したフレームデータと雑音のフレームデータに基づいて音声認識処理を行った結果、適切な音声認識結果が得られなかった場合には、制御部8は、音声認識処理に適した次候補チャンネルに該当する記憶領域をメモリアクセスしてその音声フレームデータを音声認識部7へ転送する。そして、以下同様に適切な音声認識結果が得られるまで、制御部8は順番に次候補チャンネルの音声フレームデータを音声認識部7に供給する。

【0076】

次に、かかる構成を有する本音声認識システムの動作例を図9及び図10に示すフローチャートに基づいて説明する。尚、図9は、マイクロフォン $M_1 \sim M_N$ で集音し入力データ $D_1 \sim D_N$ をフレームデータとして記憶部4に記憶するまでの動作、図10は、データ解析部6が最適の音声フレームデータと雑音データを決定する際の動作を示している。

【0077】

図9において、ステップ100では発話スイッチ9がオン操作されるまで待機し、発話スイッチ9がオン操作されるとステップ102に移行して初期化処理が行われる。この初期化処理では、制御部8内に設けられているチャンネル番号カウンタの計数値 $n$ と、フレーム番号カウンタの計数値 $m$ と、連続回数カウンタ $FC_1 \sim FC_N$ の全ての値 $F(1) \sim F(N)$ をクリアする。

【0078】

尚、上記のチャンネル番号カウンタは、マイクロフォン $M_1 \sim M_N$ の各チャンネルを計数値 $n$ によって指定するために設けられている。上記のフレーム番号カウンタは、記憶領域 $ME_1 \sim ME_N$ に備えられている各フレーム領域 $MF_1, MF_2, MF_3 \dots$ の番号（アドレス）を計数値 $m$ によって指定するために設けられている。

【0079】

上記の連続回数カウンタは $FC_1 \sim FC_N$ は、各チャンネルに対応付けて $N$ 個備えられている。すなわち、第1の連続回数カウンタ $FC_1$ は第1チャンネル、第2の連続回数カウンタ $FC_2$ は第2チャンネル、以下同様にして第 $N$ の連続回数

カウンタ  $FC_N$  は第  $N$  チャンネルに対応付けて設けられている。そして、閾値  $THD1$  より大きな  $LPC$  残差  $\varepsilon_n$  が連続して 3 フレーム以上続いたか否かと、その 3 フレーム以上続いたチャンネルを判定するために設けられている。更に、その 3 フレーム以上続いたチャンネルを発話入力のなされたチャンネルと判定するために設けられている。

【 0 0 8 0 】

次に、ステップ 1 0 4 において、記憶領域  $ME_1 \sim ME_N$  のそれぞれの第 1 番目のフレーム領域  $MF_1$  を設定する。すなわち、フレーム領域の番号  $m$  を  $m = 1$  に設定する。

【 0 0 8 1 】

次に、ステップ 1 0 6 と 1 0 8 において、マイクロフォン  $M_1 \sim M_N$  に集音動作を開始させ、それによって得られる入力データ  $D_1 \sim D_N$  を記憶領域  $ME_1 \sim ME_N$  の第 1 番目の各フレーム領域  $MF_1$  に 1 フレーム分ずつ記憶させる。

【 0 0 8 2 】

上記の 1 フレーム分ずつの入力データ  $D_1 \sim D_N$  が記憶されると、次にステップ 1 1 0 において、第 1 ( $n = 1$ ) のチャンネルに該当する記憶領域  $ME_1$  を指定した後、ステップ 1 1 2 において記憶領域  $ME_1$  の第 1 番目 ( $m = 1$ ) のフレーム領域  $MF_1$  に記憶されたフレームデータの  $LPC$  残差  $\varepsilon_n$  ( $n = 1$ ) を演算する。

【 0 0 8 3 】

次に、ステップ 1 1 4 において、その  $LPC$  残差  $\varepsilon_n$  と閾値  $THD1$  とを比較する。 $\varepsilon_n \geq THD1$  のときにはステップ 1 1 6 に移行して第 1 チャンネルに該当する連続回数カウンタ  $FC_1$  の値  $F(1)$  をインクリメント (1 を加算) する。 $\varepsilon_n < THD1$  のときにはステップ 1 1 8 に移行して連続回数カウンタ  $FC_1$  の値  $F(1)$  をクリアする。

【 0 0 8 4 】

したがって、 $\varepsilon_n \geq THD1$  となると、連続回数カウンタ  $FC_1$  の値  $F(1)$  は 1 になるので、この値  $F(1)$  は、第 1 チャンネルのマイクロフォン  $M_1$  に対し 1 フレーム分の発話入力となされたことを示すことになる。

## 【 0 0 8 5 】

一方、 $\varepsilon_n < \text{THD1}$  となると、連続回数カウンタ  $\text{FC}_1$  の値  $F(1)$  が 0 にクリアされるので、この値  $F(1)$  は、第 1 チャンネルのマイクロフォン  $M_1$  に対し発話入力がなされていないことを示すことになる。

## 【 0 0 8 6 】

次に、ステップ 1 2 0 において  $n = N$  か調べることにより、全てのチャンネルにおける LPC 残差  $\varepsilon_n$  を演算したか判断する。 $n = N$  でなければステップ 1 2 2 に移行して、 $n = n + 1$  とすることにより、次のチャンネルを設定してステップ 1 1 2 からの処理を繰り返す。すなわち、ステップ 1 1 2 ~ 1 2 2 の処理を  $N$  回繰り返すことで、記憶領域  $\text{ME}_1 \sim \text{ME}_N$  の各フレーム領域  $\text{MF}_1$  に記憶されたフレームデータの LPC 残差  $\varepsilon_n$  と閾値  $\text{THD1}$  とを比較する。そして、LPC 残差  $\varepsilon_n$  が閾値  $\text{THD1}$  以上となったときには、そのチャンネル番号  $n$  に該当する連続回数カウンタ  $\text{FC}_n$  の値  $F(n)$  をインクリメントする。

## 【 0 0 8 7 】

次に、上記ステップ 1 2 0 において  $n = N$  となると、全チャンネルについての処理が完了したと判断してステップ 1 2 4 に移行する。

## 【 0 0 8 8 】

ステップ 1 2 4 では、連続回数カウンタ  $\text{FC}_1 \sim \text{FC}_N$  の値  $F(1) \sim F(N)$  のうち、値が 3 以上となったものがあるか判断する。値が 3 以上となったものが無ければ、すなわち、何れの値  $F(1) \sim F(N)$  も 2 以下であればステップ 1 2 6 に移行する。

## 【 0 0 8 9 】

ステップ 1 2 6 では、 $m = m + 1$  とすることにより記憶領域  $\text{ME}_1 \sim \text{ME}_N$  の第 2 番目の各フレーム領域  $\text{MF}_2$  を設定する。そして、ステップ 1 0 6 ~ 1 2 4 の処理が繰り返される。

## 【 0 0 9 0 】

これにより、各フレーム領域  $\text{MF}_2$  に入力データを記憶させ（ステップ 1 0 6 , 1 0 8）、更に各フレーム領域  $\text{MF}_2$  に記憶された各フレームデータの LPC 残差  $\varepsilon_2$  を閾値  $\text{THD1}$  と比較し（ステップ 1 1 0 ~ 1 1 4）、その比較結果に

基づいて各連続回数カウンタ  $FC_1 \sim FC_N$  の各値  $F(1) \sim F(N)$  をインクリメント又はクリアすることになる。

#### 【0091】

そして、再びステップ 1 2 4 において、連続回数カウンタ  $FC_1 \sim FC_N$  の値  $F(1) \sim F(N)$  のうち、値が 3 以上となったものがあるか判断し、無ければステップ 1 3 0 に移行して、更に  $m = m + 1$  とすることにより記憶領域  $ME_1 \sim ME_N$  の次の各フレーム領域  $MF_3$  を設定してステップ 1 0 6 ～ 1 2 4 の処理を繰り返す。

#### 【0092】

こうしてステップ 1 0 6 ～ 1 2 6 の処理を繰り返し、ステップ 1 2 4 において、連続回数カウンタ  $FC_1 \sim FC_N$  の値  $F(1) \sim F(N)$  のうち、少なくとも 1 つの連続回数カウンタの値が 3 以上となると、ステップ 1 2 8 の処理に移行する。

#### 【0093】

つまり、ステップ 1 2 4 では、連続回数カウンタ  $FC_1 \sim FC_N$  の値  $F(1) \sim F(N)$  を調べ、閾値  $THD_1$  より大きな LPC 残差  $\varepsilon_n$  が連続して 3 フレーム以上続いた場合に限り、そのチャンネルに該当する記憶領域に記憶されたフレームデータを音声フレームデータであると判断し決定する。

#### 【0094】

次に、ステップ 1 2 8 では、上記の閾値  $THD_1$  より大きな LPC 残差  $\varepsilon_n$  が連続して 3 フレーム以上続いたと判断された連続解するカウンタの値が更に 5 に達したか否か判断し、値が 5 に達していなければステップ 1 2 6 の処理を行って、再びステップ 1 0 6 ～ 1 2 8 の処理を繰り返す。

#### 【0095】

すなわち、或るチャンネル  $n$  に該当する連続回数カウンタ  $FC_n$  の値  $F(n)$  が 3 になったときに、残余のチャンネルに該当する連続回数カウンタの値が 1 又は 2 になっている場合がある。こうした場合には、残余のチャンネルに該当する記憶領域に記憶されたフレームデータも音声フレームデータである可能性がある。

#### 【0096】

そこで、上記ステップ 1 0 6 ～ 1 2 8 の処理を更に 2 回繰り返すことにより、残余のチャンネルに該当する記憶領域に記憶されたフレームデータが音声フレー

ムであるか否か確認することとしている。

【 0 0 9 7 】

次に、ステップ 1 2 8 の判断が「Y E S」となると、ステップ 1 3 0 に移行し、音声フレームデータが記憶された記憶領域と雑音フレームデータが記憶された記憶領域との情報を有する音声検出データ D C T 1 を制御部 8 へ転送した後、図 1 0 に示す処理に移行する。

【 0 0 9 8 】

図 1 0 の処理に移行すると、まずステップ 2 0 0 において、各チャンネルの平均音声パワー  $P(n)$  と平均雑音パワー  $N P(n)$  及び平均  $S/N$  値  $S N(n)$  を演算する。次に、ステップ 2 0 2 において、図 7 に示した音声条件テーブル 6 f に基づいて音声認識に適した音声フレームデータの候補を決定する。更に、ステップ 2 0 4 において、図 8 に示した雑音選択テーブル 6 g に基づいて音声認識に適した雑音フレームデータを決定する。

【 0 0 9 9 】

そして、ステップ 2 0 6 において、データ解析部 6 から制御部 8 へ、音声認識に適した音声フレームデータの候補を示す音声候補データ D C T 2 と雑音フレームデータを示す雑音候補データ D C T 3 が転送される。つまり、音声候補データ D C T 2 と雑音候補データ D C T 3 によって、音声認識に適した音声フレームデータの候補と、各候補に対応する音声認識に適した雑音フレームデータとが関連付けられて制御部 8 へ通知される。

【 0 1 0 0 】

次に、ステップ 2 0 8 において、音声認識部 7 が最も音声認識に適した音声フレームデータと雑音フレームデータを記憶部 4 から読み出して音声認識処理を行い、適切な音声認識結果が得られた場合には一連の音声認識処理を終了する。

【 0 1 0 1 】

一方、適切な音声認識結果が得られなかった場合には、ステップ 2 1 2 において次の候補の音声フレームデータと雑音フレームデータがあるか否か調べた上で、次の候補の音声フレームデータと雑音フレームデータとを記憶部 4 から読み出し、再びステップ 2 0 8 からの処理を繰り返す。また、再度の音声認識処理を行

っても適切な音声認識結果が得られなかった場合には、同様に次の候補の音声フレームデータと雑音フレームデータを読み出して音声認識処理を行い、適切な音声認識結果が得られるまでステップ 2 0 8 ~ 2 1 2 の処理を繰り返す。

#### 【 0 1 0 2 】

このように本実施形態によれば、複数個の音声入力用のマイクロフォン  $M_1 \sim M_n$  を車室内に配置し、これらのマイクロフォン  $M_1 \sim M_n$  で集音されることで生じる音声フレームデータと雑音フレームデータの中から、音声認識に適した音声フレームデータと雑音フレームデータを自動的に抽出して音声認識処理を行うようにしたので、従来の一人の発話者を対象とする音声認識システムと較べて、複数の発話者（搭乗者）に対し極めて優れた利便性を提供することができる。

#### 【 0 1 0 3 】

例えば、複数の搭乗者のうちの一人が、あるマイクロフォン（例えば  $M_1$ ）に対して所望の発話を行った場合に、他のマイクロフォン（ $M_2 \sim M_N$ ）にもその発話音声が入り込んで集音され、その結果どのマイクロフォンで本当の発話音声が集音されたか解らなくなる状況が一般的には想定されるが、本実施形態によれば、図 7 及び図 8 に示した音声条件テーブル 6 f と雑音選択テーブル 6 g とを用いて、音声認識に適した音声フレームデータと雑音フレームデータを自動的に抽出して音声認識処理を行うようにしたので、発話した搭乗者とその搭乗者に近い位置のマイクロフォン（例えば  $M_1$ ）とが極めて高い確率で対応付けられることになる。

#### 【 0 1 0 4 】

このため、本音声認識システムは、音声操作を行おうとする搭乗者を自動的に特定してその発話音声を最適の（近い位置の）マイクロフォンで集音することになるため、音声認識率を向上させることができる。一方、搭乗者の側から見ると、特別な操作をしなくとも、単に発話するだけで自分の発話指令が適切なマイクロフォンを通して受け入れられることになるため、極めて利便性の高い音声認識システムの構築が可能となっている。

#### 【 0 1 0 5 】

更にまた、音声操作を行おうとしていない 1 又は複数の搭乗者が会話等を行っ



ている状況で、ある一人が音声操作のための発話を行うような場合でも、図7及び図8に示した音声条件テーブル6fと雑音選択テーブル6gとを用いて、音声認識に適した音声フレームデータと雑音フレームデータを自動的に抽出して音声認識処理を行うことで、音声操作を行おうとしていない搭乗者の会話等は雑音として判定されることになって除外される。このため、周囲の会話等の影響を受けない音声認識システムを実現することができ、ひいては、極めて利便性の高い音声認識システムの構築が可能となっている。

## 【0106】

尚、本実施形態は、図1に示したように、発話スイッチ9を1個だけ設け、例えば運転手がこの発話スイッチ9をオン操作するようにしたが、本発明は、かかる構成に限定されるものではない。例えば、図11のブロック図に示すように、複数のマイクロフォン $M_1 \sim M_N$ のそれぞれに発話スイッチ $TK_1 \sim TK_N$ を設けておき、いずれかの発話スイッチがオン操作されると、制御部3が、そのオン操作された発話スイッチに対応するマイクロフォンに発話入力を行わせ、オン操作されない発話スイッチに対応する残余のマイクロフォンでは車室内の雑音が集音されたと判断するようにしてもよい。

## 【0107】

かかる構成によると、発話音声を集音したマイクロフォンと、雑音を集音したマイクロホンとを音声認識処理の前に予め特定することが可能となるため、最も音声認識に適した音声データと雑音データを容易に決定するための処理時間を短縮化することができる。

## 【0108】

また、図1の構成と図11の構成との折衷案の構成としてもよい。すなわち、マイクロフォン $M_1 \sim M_N$ の数よりも少ない個数の発話スイッチを車室内の適宜の位置に配置し、何れかの発話スイッチがオン操作されると、制御部8がこれを検出して音声認識処理を開始するようにしてもよい。この場合には、各発話スイッチとマイクロフォン $M_1 \sim M_N$ が完全に一対一に対応付けられないため、図1に示した構成によって音声認識処理が行われることになるが、発話音声を集音したマイクロフォンと雑音を集音したマイクロホンとを音声認識処理の前に予め特定す

ることが可能となるため、音声認識に適した音声フレームデータと雑音フレームデータを決定するための処理時間を短縮化することができる。

【0 1 0 9】

更に、マイクロフォン $M_1 \sim M_N$ より少ない数の発話スイッチを設けて図 1 の構成を適用する場合に、各発話スイッチをマイクロフォンの設置範囲として決めておき、それぞれの設置範囲内に属する 1 又は複数個のマイクロフォンを、何れの発話スイッチがオン操作されたかで予め特定するようにしてもよい。かかる構成によれば、予め特定された 1 又は複数個の音声フレームデータと雑音フレームデータから音声認識処理に適したものを抽出すればよいので、処理時間の短縮化が可能となる。

【0 1 1 0】

また、本実施形態では、車載用電子機器に適用する音声認識システムについて説明したが、本発明の音声認識システムは、汎用のマイクロコンピュータシステムや所謂ワードプロセッサ等の電子機器に適用して、発話による文書入力を行うことも可能である。

【0 1 1 1】

【発明の効果】

以上説明したように本発明によれば、発話者が所望の発話をする、と、複数個の音検知手段から出力される各入力信号の中から、音声認識を行うのに適した音声信号と雑音信号が自動的に決定され、その決定された音声信号と雑音信号に基づいて音声認識が行われるようにしたので、発話者にとって特定の音声入力手段に対して発話するという意識を持つことなく適切な音声認識がなされることになり、利便性の高い音声認識システムを提供することができる。

【図面の簡単な説明】

【図 1】

本実施形態に係る音声認識システムの構成を示すブロック図である。

【図 2】

マイクロフォンの配置態様の一例を示す説明図である。

【図 3】

マイクロフォンの他の配置態様の一例を示す説明図である。

【図 4】

マルチプレクサとデマルチプレクサ及び記憶部の構成を示すブロック図である。

【図 5】

入力信号をサンプリングして記憶部に記憶させる際のタイミングを説明するためのタイミングチャートである。

【図 6】

平均音声パワーと平均雑音パワー及び平均 S / N 値の演算方法を説明するための説明図である。

【図 7】

音声条件テーブルの構成を示す説明図である。

【図 8】

雑音選択レーブルの構成を示す説明図である。

【図 9】

本実施形態に係る音声認識システムの動作を説明するためのフローチャートである。

【図 1 0】

本実施形態に係る音声認識システムの動作を更に説明するためのフローチャートである。

【図 1 1】

本実施形態に係る音声認識システムの変形例の構成を示すブロック図である。

【符号の説明】

- 1 … マルチプレクサ
- 2 … A / D 変換器
- 3 … デマルチプレクサ
- 4 … 記憶部
- 5 … 発話音声検出部

6 …データ解析部

6 a …最適音声決定部

6 b …雑音決定部

6 c …平均 S / N 値演算部

6 d …平均音声パワー演算部

6 e …平均雑音パワー演算部

6 f …音声条件テーブル

6 g …雑音選択テーブル

7 …音声認識部

8 …制御部

9,  $TK_1 \sim TK_N$  …発話スイッチ

$M_1 \sim M_N$  …マイクロフォン

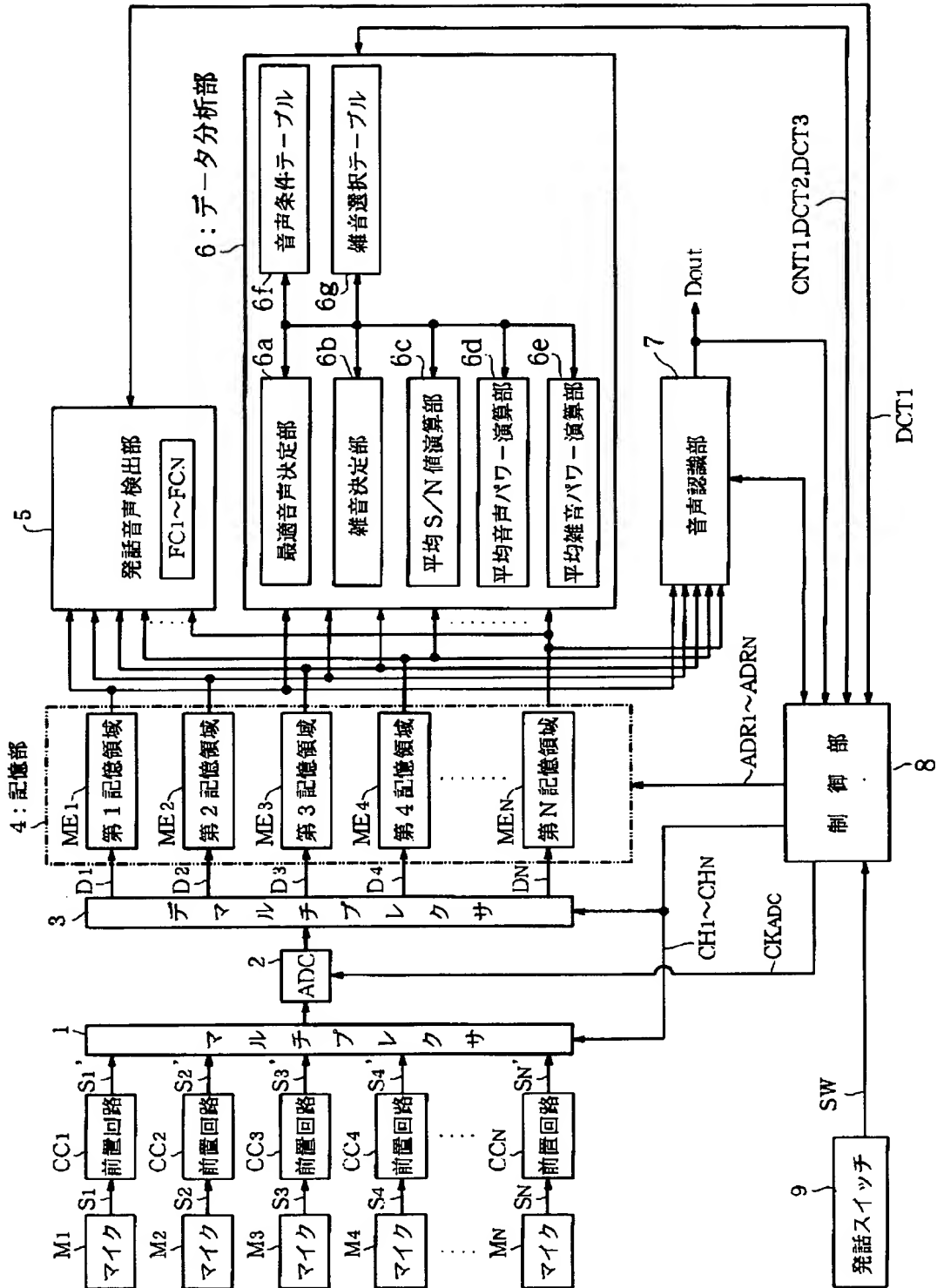
$ME_1 \sim ME_N$  …記憶領域

$MF_1 \sim MF_N$  …フレーム領域

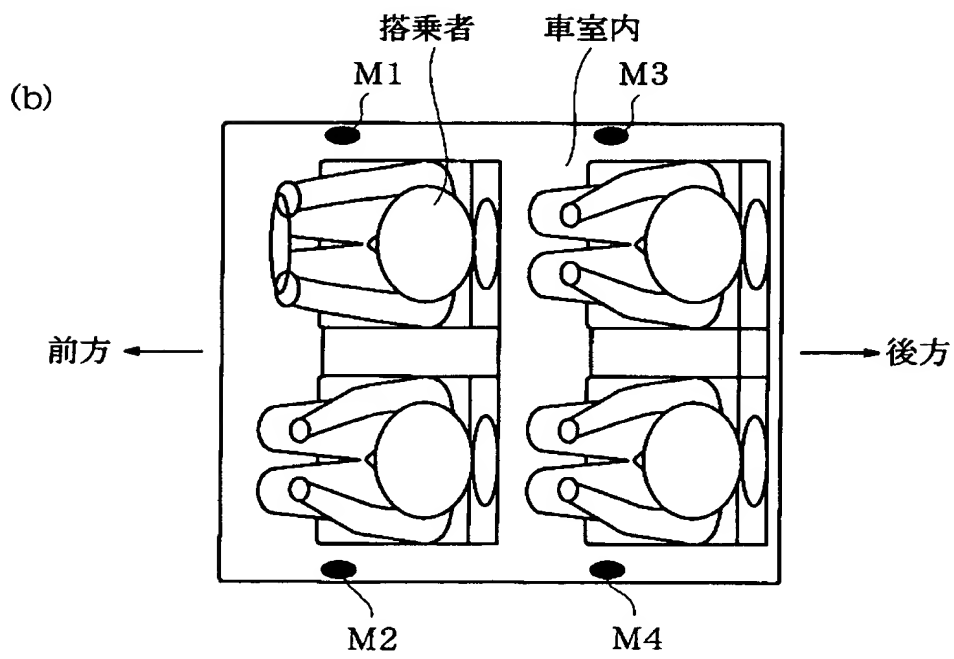
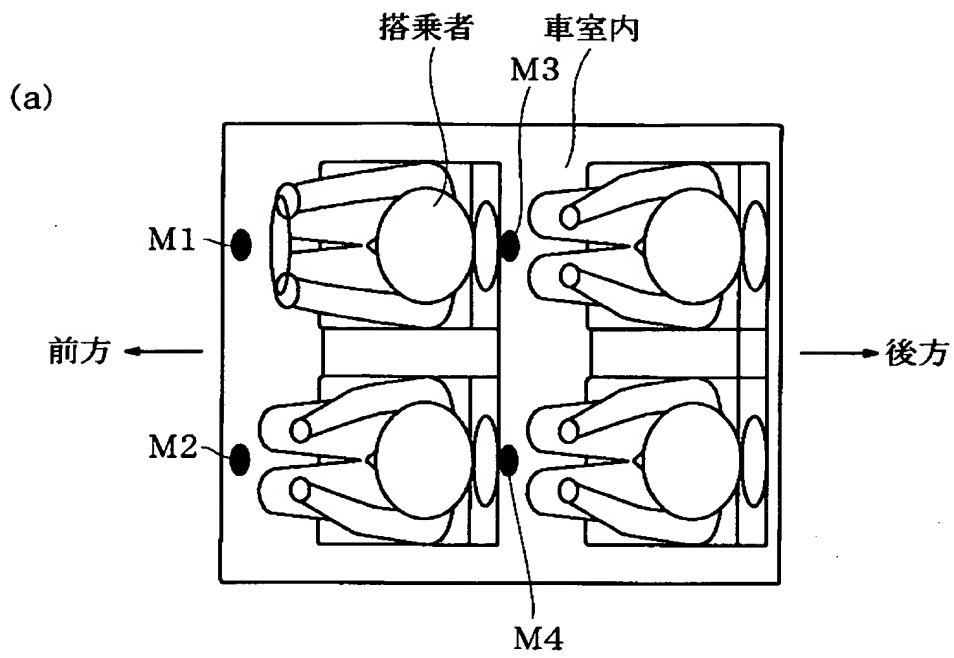
$AS_1 \sim AS_N, AW_1 \sim AW_N$  …アナログスイッチ

【書類名】 図面

【図 1】

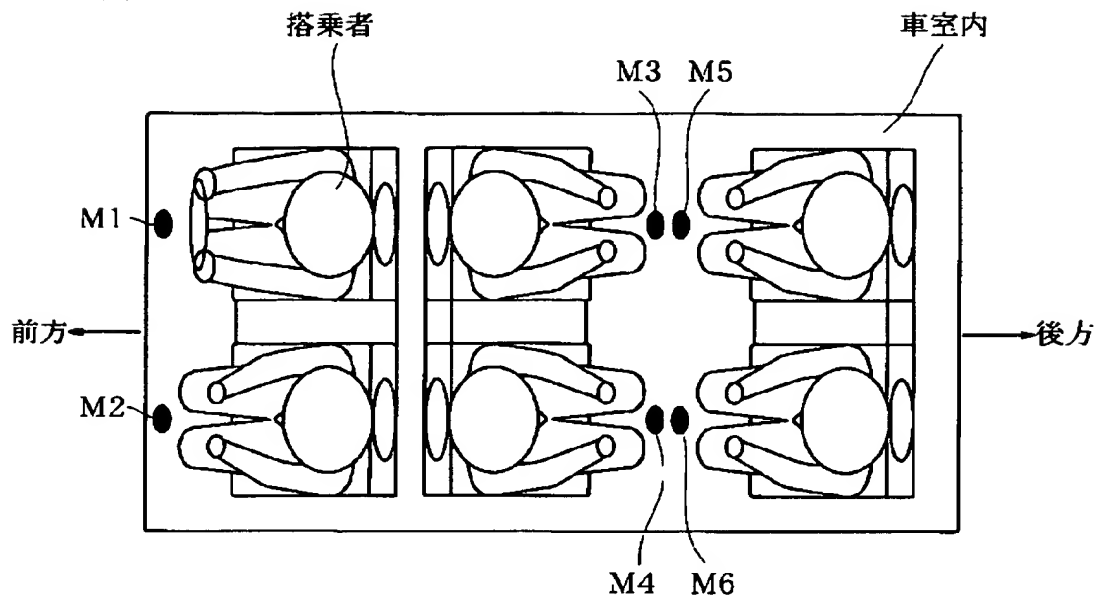


【図 2】

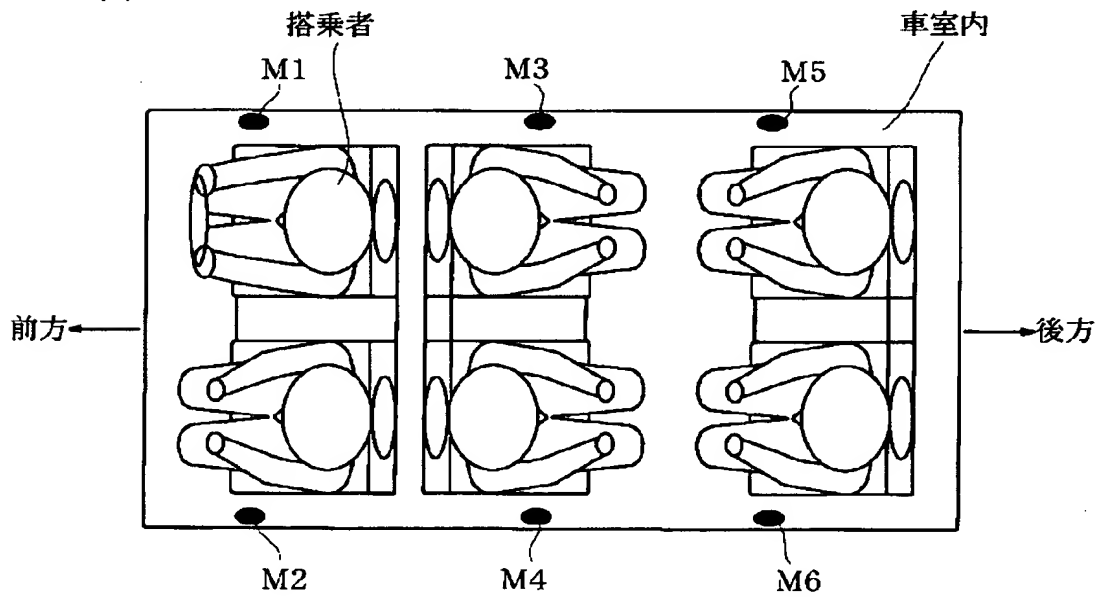


【図 3】

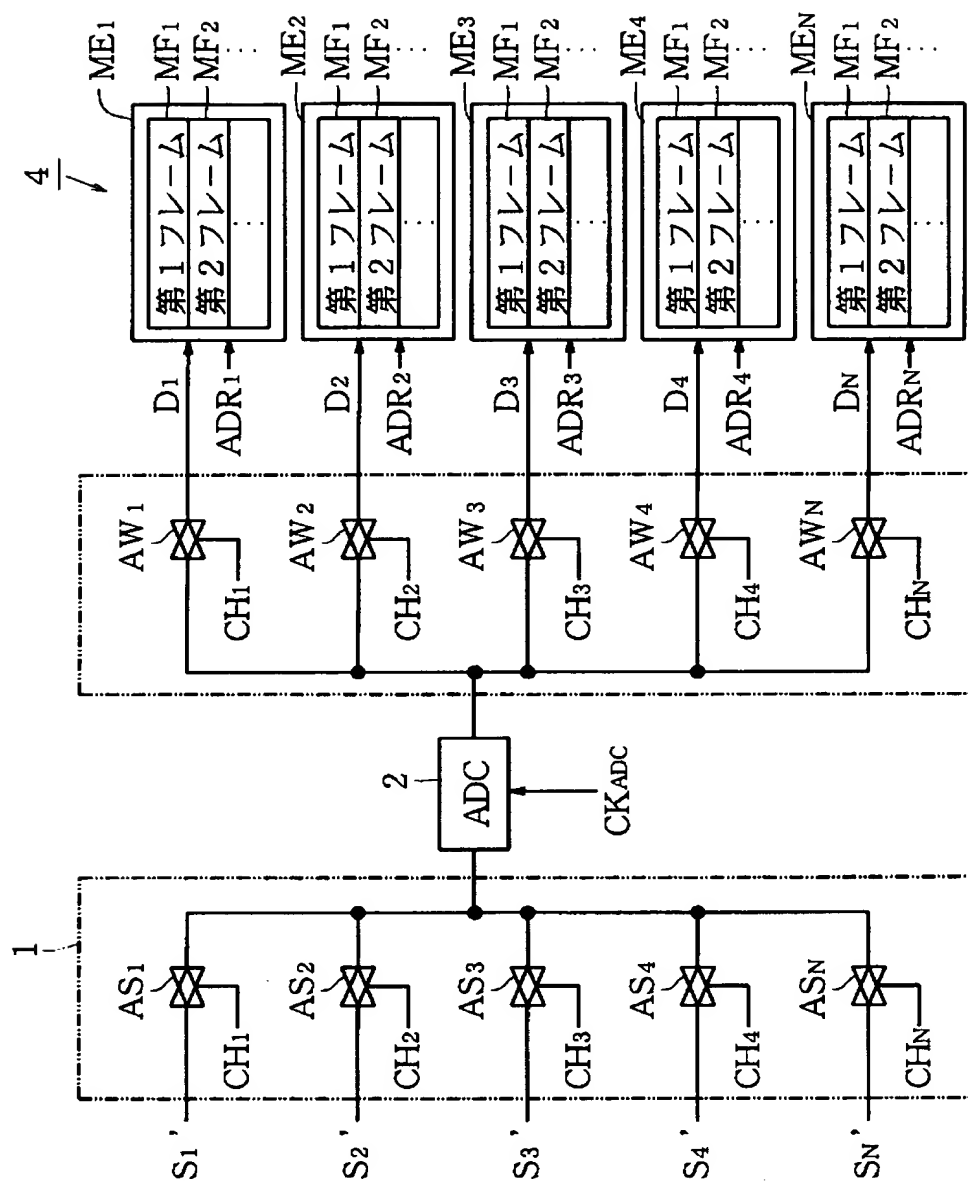
(a)



(b)

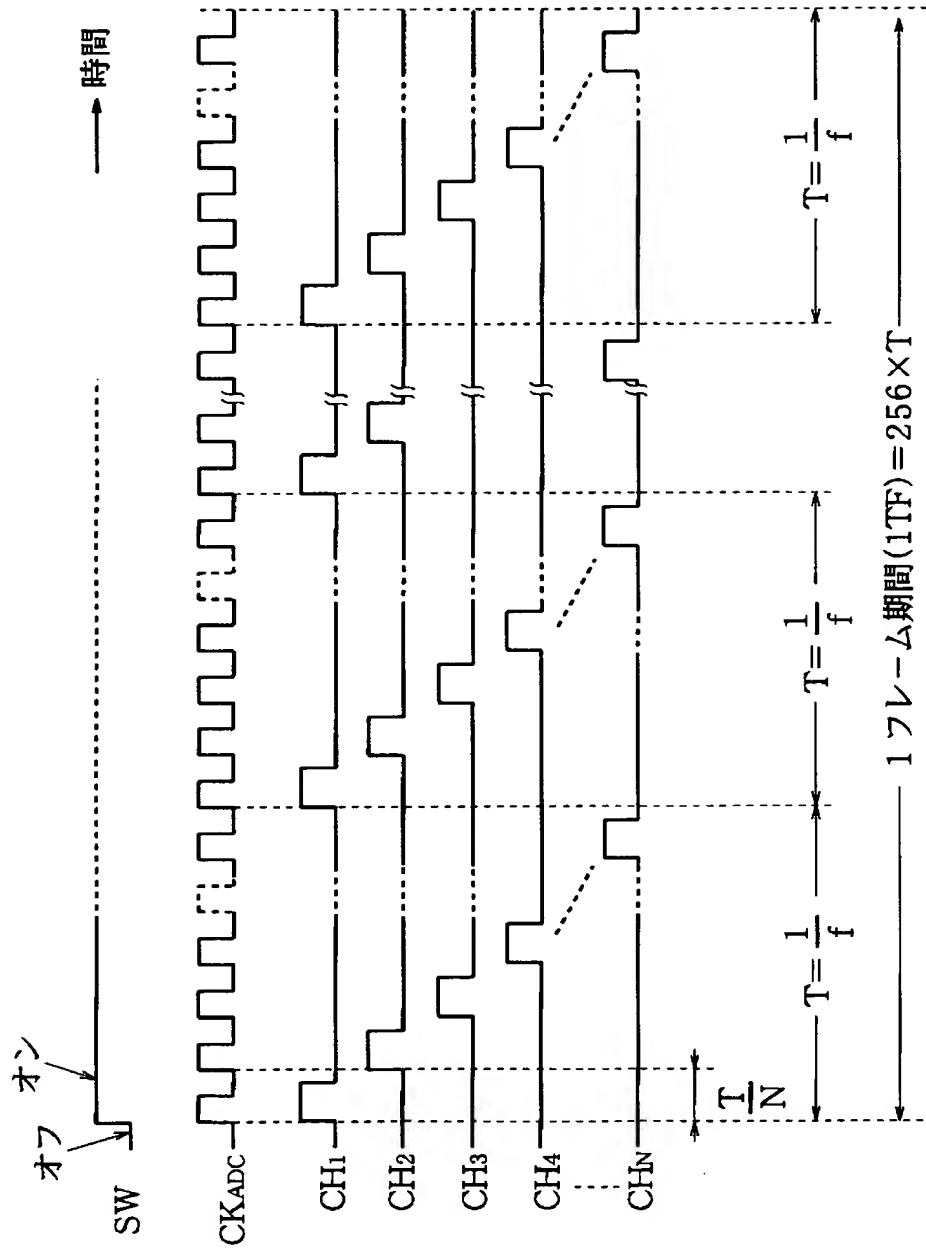


【図 4】



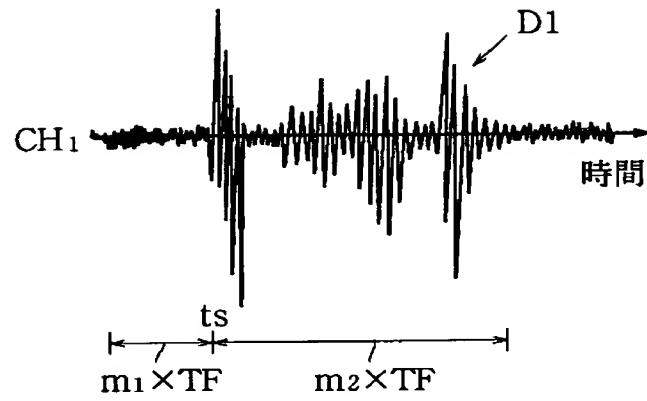


【図 5】

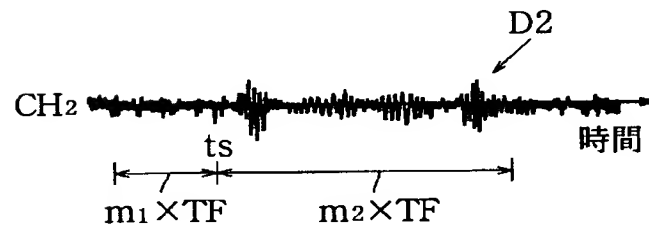


【図 6】

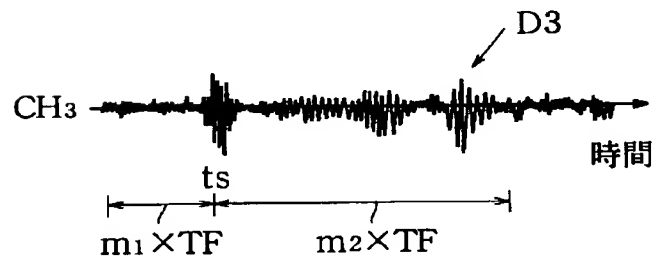
(a)



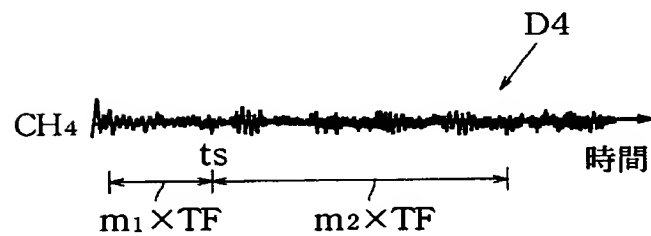
(b)



(c)



(d)



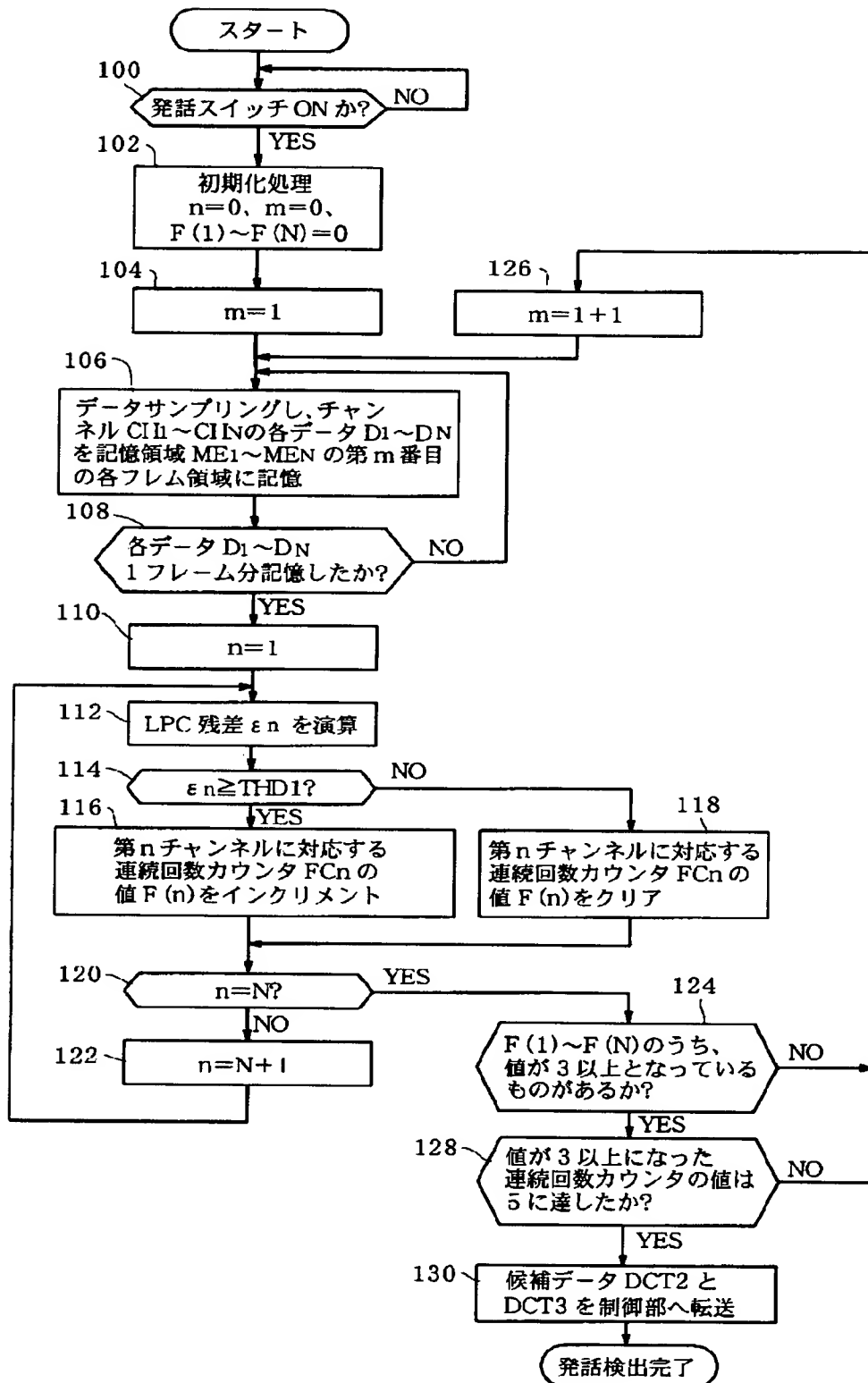
【図 7】

等級	平均 S/N	平均音声パワー
Rnk0	最大値	最大値
Rnk1	THD2 以上	THD3 以上
Rnk2	THD2 以上	THD3 未満
Rnk3	THD2 未満	THD3 以上
Rnk4	THD2 未満	THD3 未満

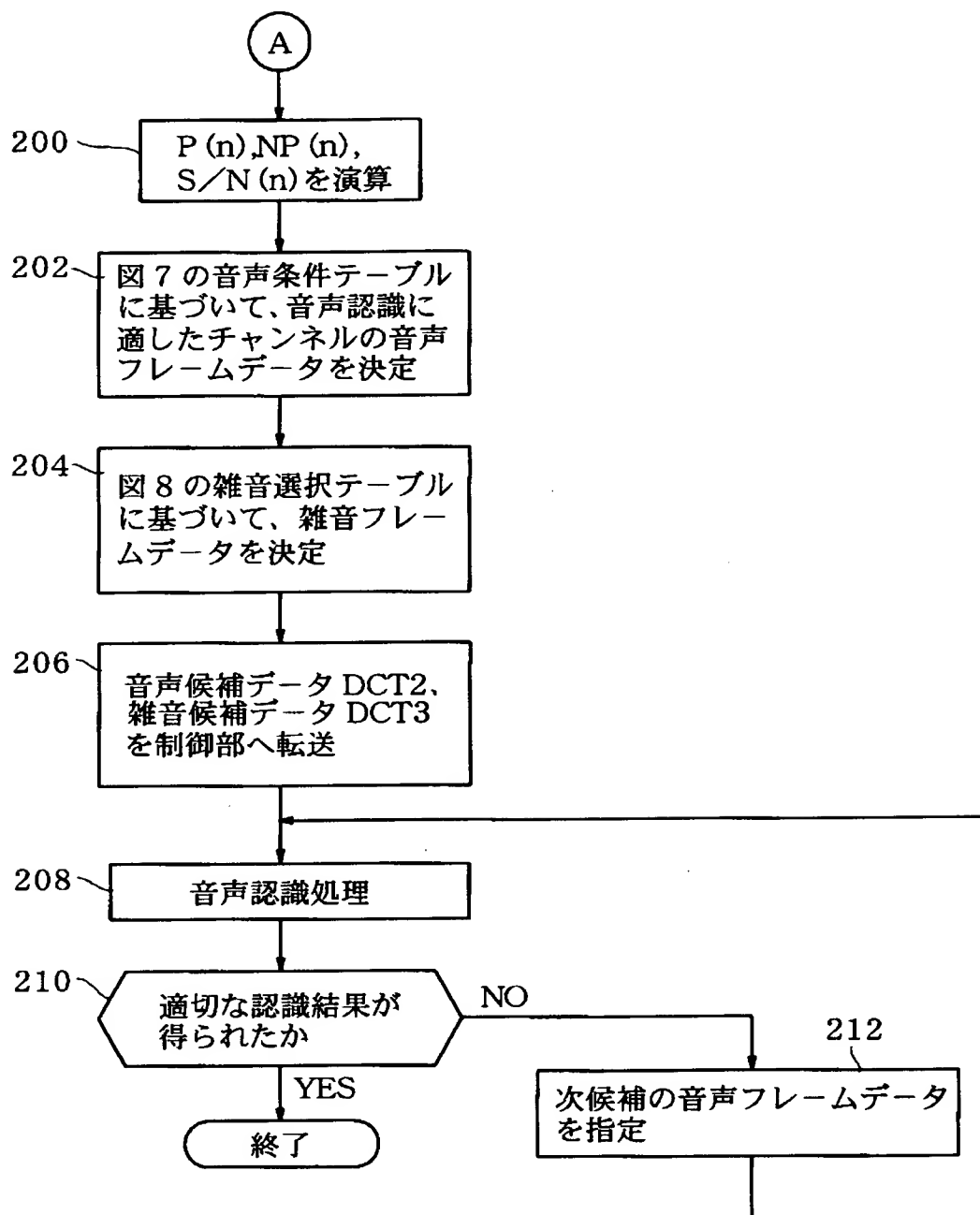
【図 8】

	CH1	CH2	CH3	CH4	……	雑音のチャンネル
場合 1	Rnk0	Rnk1	Rnk2	Rnk1	……	CH3
場合 2	Rnk0	Rnk1	Rnk1	Rnk2	……	CH4
場合 3	Rnk0	Rnk1	Rnk1	Rnk3	……	CH4
場合 4	Rnk0	Rnk1	Rnk1	Rnk4	……	CH4
場合 5	Rnk0	Rnk1	Rnk2	Rnk1	……	CH3
場合 6	Rnk0	Rnk1	Rnk2	Rnk2	……	CH3
場合 7	Rnk0	Rnk1	Rnk2	Rnk3	……	CH3
場合 8	Rnk0	Rnk1	Rnk2	Rnk4	……	CH3
場合 9	Rnk0	Rnk1	Rnk3	Rnk1	……	CH3

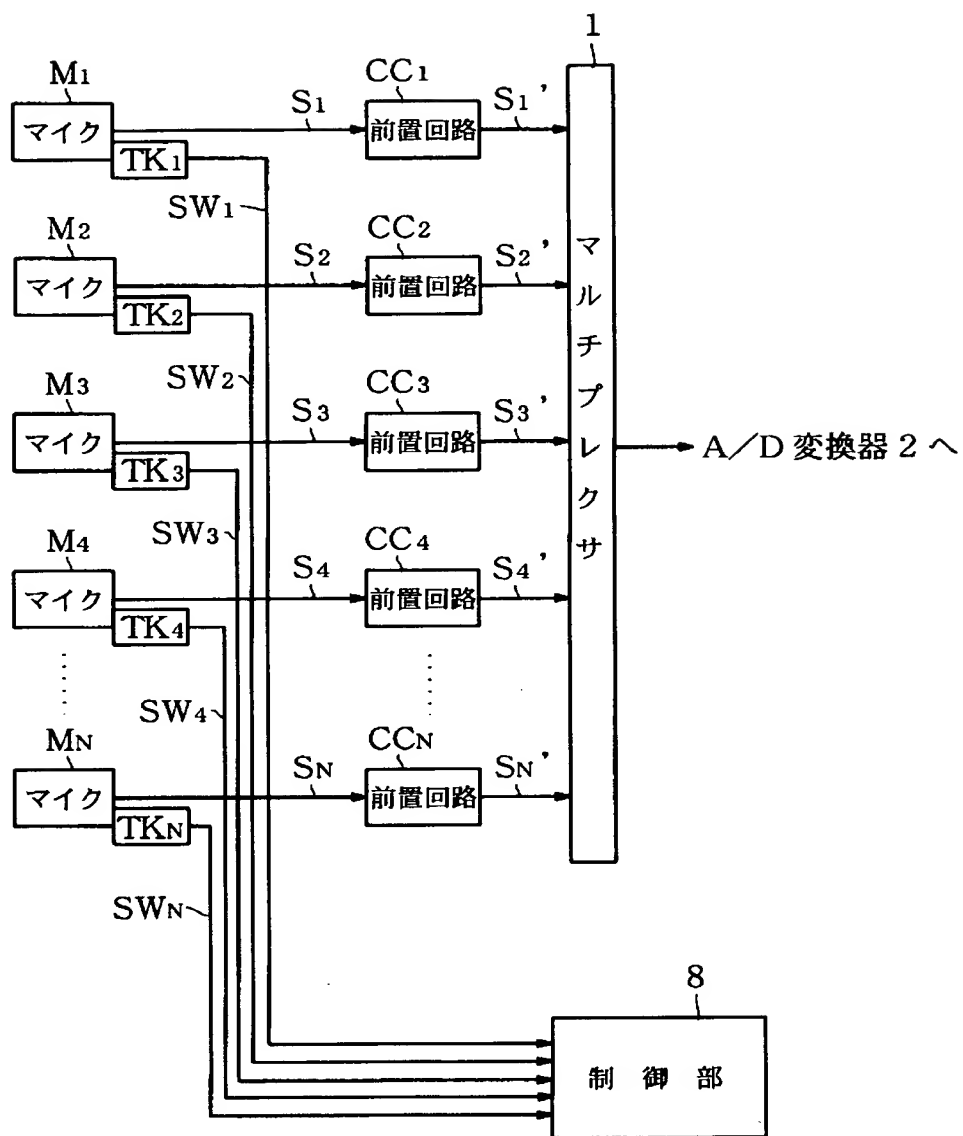
【図 9】



【図 1 0】



【図 1 1】



【書類名】 要約書

【要約】

【課題】 利便性の高い音声認識システムを提供する

【解決手段】 複数個のマイクロフォン $M_1 \sim M_N$ を備え、マイクロフォン $M_1 \sim M_N$ から出力される入力信号 $S_1 \sim S_N$ を入力データ $D_1 \sim D_N$ に変換して記憶部4の各記憶領域 $ME_1 \sim ME_N$ に所定のフレーム単位で記憶させる。各フレーム単位の入力データ $D_1 \sim D_N$ の各平均 $S/N$ 値と平均音声パワーとを演算し、平均 $S/N$ 値と平均音声パワーが所定の閾値より大きくなる入力データを音声認識に適した音声データ、平均音声パワーと平均雑音パワーが最も小さくなる入力データを雑音データと決定し、これらの音声データと雑音データに基づいて音声認識を行う。

【選択図】 図1

出 願 人 履 歴 情 報

識別番号 [000005016]

1. 変更年月日	1990年 8月31日
[変更理由]	新規登録
住 所	東京都目黒区目黒1丁目4番1号
氏 名	パイオニア株式会社